

TESTING THRESHOLD COINTEGRATION IN WAGNER'S LAW: THE ROLE OF MILITARY SPENDING

Maddalena Cavicchioli^a Barbara Pistoresi^b

Abstract. *This paper uses historical data since mid-19th century to test the validity of Wagner's Law for the Italian economy. Unlike the previous studies, we accommodate possible nonlinear asymmetric effects of government spending and GDP toward their long-run equilibrium. Our results show the presence of a threshold cointegrating relationship between the two variables with significantly different error correction adjustments in normal and extreme regimes. Particularly, we find the validity of Wagner's Law from 1862 to 2009, only when we take into account strong nonlinear responses of government spending during the WWI and WWII period. Robustness checks recognize nonlinear behaviour of government expenditure driven by military spending.*

Keywords: Time series; Nonlinearity; Threshold Vector Error Correction; Public spending; Economic growth; Wagner's Law.

JEL Classification: C01; C05; C34; H1; H5.

^a*Corresponding Author:* Department of Economics "Marco Biagi", Viale Berengario 51, 41121, Modena (Italy). Office Phone No: +39 059 2056846. E-mail: <maddalena.cavicchioli@unimore.it>

^bDepartment of Economics "Marco Biagi" and ReCent, Viale Berengario 51, 41121, Modena (Italy). Office Phone No: +39 059 2056881. E-mail: <barbara.pistoresi@unimore.it>

1. Introduction

The long-run relationship between the size of public sector and economic growth remains an important accepted stylized fact in the literature of public economics (e.g. Atkinson and Stiglitz (1980)). A simple explanation for the long-run determination of public spending was proposed by Wagner (1883) and it is known as Wagner's Law (henceforth WL). It states that the growth of government expenditure is a consequence of the expansion of the state produced by a country's social and economic development. The urbanization and greater division of labour accompanying industrialization require, for example, more government regulation and higher expenditure on contractual enforcement and law and order. Other reasons are the growing need to finance large-scale investments with public good characteristics (i.e., infrastructures) and the supposed superior income-elasticity of publicly provided goods and services, such as education, welfare, but also national security or defence.

A large number of studies focus on an empirical assessment of WL from different perspective and with different techniques. See, for recent reviews, Durevall and Henrekson (2011) and Kuckruck (2014). Most of the empirical literature focuses on developed or developing economies over relatively short time spans, generally starting from the 1960s. The majority compare the results for industrialized and emerging economies in order to make a statement about the relation between development level and WL, although there are significant differences between a modern state in the 19th century and recent developing economies, in terms of culture, institutions and conception of the state activity.

By contrast, the analysis of WL in a long-run perspective, on a single country or countries with similar social, economic and political conditions has attracted much less attention. Few studies analyse very long time spans and generally reject WL. Henrekson (1993) and Bohl (1996) find no support for WL in Sweden from 1861 to 1990 and in the United Kingdom from 1870 to 1995, respectively; Ghate and Zak (2002) do not find any empirical evidence in the United States from 1929 to 2000; Durevall and Henrekson (2011) find direct evidence in favor of WL only for Sweden and the United Kingdom in a time period from around 1860 to 1970. There are, however, shorter time spans during which WL holds. For example, Oxley (1994) for the United Kingdom, Thornton (1999) for Denmark, Germany, Italy, Norway, Sweden, and the United Kingdom, and Durevall and Henrekson (2011) for Sweden and the United Kingdom confirm the validity of WL in a period of 50 years preceding World War I. Recently, Kuckuck (2014) examines UK, Denmark, Sweden, Finland and Italy, finding that a long-run equilibrium between public spending and economic growth exists but WL has a higher validity during

early stages of development.

The econometric techniques to study this topic concern: (i) explicit focus on structural breaks to evaluate shifts in the long-run; (ii) tests for asymmetric adjustment in the short-run differentiating positive or negative deviations from the trend; (iii) preliminary detection of potential outliers before proceeding to cointegration analysis (see Durevall and Henrekson (2011) for a recent review). Here we aim at evaluating the presence of endogenous nonlinearities within the analysis of the long-run relationship between the variables of interest.

Particularly, the goal of our paper is to contribute to the analysis of WL for the Italian economy, from its political unification in 1861 to 2009. Italy is a good case for study because it is a late-comer which caught up with industrialization in the late 19th century and then exhibited an excellent economic performance that enabled it to join the G7 group in the 1970s. We consider the long span as the correct framework to test WL since it enables study of the evolution of government expenditure in response to country's social and economic progress, i.e. to the changes in demands and needs of the society. However, during 140 years, Italy has experienced a number of economic and socio-political potential sources of nonlinearities in the data: the WWI and WWII, as well as the Great Depression, and the socio-political turmoils in the post-war period. Such events might be causes of different asymmetric responses in government spending to variations in national income. If this is the case, not taking into account those features of the data might induce biased empirical results and misleading conclusions. Actually, we find the validity of WL over the period 1862-2009, only when we take into consideration strong asymmetric responses of government spending during the WWI and WWII period. Robustness checks clearly recognize nonlinear behaviour of government spending driven by the explosion of military expenditure.

The main contributions of the paper to the existing literature are the following. Firstly, unlike previous studies, we consider nonlinear cointegration to validate WL. We show that for the Italian case strong support to WL is found when nonlinearities are modelled. Hence, the presence of asymmetric adjustments in the response of government spending may explain why the bulk of empirical evidence concerning WL is inconclusive. Secondly, we apply the methodology by Hansen and Seo (2002) to incorporate the possibility of threshold effects in the cointegrating relationship, so that we nest linear cointegration allowing the potential existence of one or more regimes. Finally, our paper also differs from existing studies on Italy because it relies on a new dataset recently provided by Italy's State General Accounting Department containing data on total government expenditure. Concerning the

GDP series, our dataset also makes use of the most recent series of Italy's national accounts recently provided by the Bank of Italy.

The paper is organized as follows. Section 2 provides a theoretical framework for the empirical analysis. Moreover, it outlines the nonlinear cointegration model proposed by Hansen and Seo (2002). In Section 3 we describe the data and comment on some stylized facts. Section 4 presents the empirical results. In Section 5 we deal with some robustness checks. Section 6 concludes.

2. The Econometric Framework

In order to test WL, the literature uses different functional forms linking public spending and national income. This paper uses the following specification

$$g_t = \alpha + \theta y_t \quad (1)$$

where g is the logarithm of total government expenditure in nominal terms as a share of nominal GDP, and y is the logarithm of real per capita GDP. The above formulation is probably the most common and the majority of other models in the literature are simple reformulations of it (see Durevall *et al.* 2011). Contrary to the alternative specification that consists of total government expenditure and GDP, it has the advantage, by using real per capita GDP, of better considering a nation's prosperity and the spending capacity of its citizens. Moreover, the use in Equation (1) of total government expenditure as a share of GDP instead of real total government expenditure considers the possibility that differences in productivity growth of government and private sector production lead to an increase in government spending due to "Baumol's disease"¹. Now, Equation (1) models the evolution of the demand for public goods and services in the long-run through the coefficient θ that according to WL should be greater than zero. In this case, the government expenditure increases faster than GDP, i.e., government expenditure is income-elastic, or in other words, it is a superior good. Cointegration and Granger causation offer an econometric framework to estimate Equation (1). Note that WL requires that if a long-run relationship exists, i.e., that g and y are cointegrated, also g must Granger-cause y and not vice versa.

However, if previous works have massively employed cointegration analysis, only some of them have adequately addressed the issue of regime change and,

¹"Baumol's disease" involves a rise of wages in jobs - i.e., in the government sector - that have experienced no increase of labour productivity in response to rising wages in other jobs - i.e., in private sector production - which experienced such labour productivity growth. The rise of wages in jobs without productivity gains is caused by the requirement to compete for employees with jobs that experienced gains and hence can naturally pay higher salaries (Baumol (1967)).

generally speaking, of any asymmetric movements in the relationship between the variables. If this is the case, it might be of some importance to allow in the same cointegration framework the possibility of linear and nonlinear effects. This could be potentially meaningful to capture the underlying dynamics of the data. Note that in the case of non-cointegrating variables, threshold vector autoregression models (TVAR) (see Tsay (1998) or Ferraresi et al. (2015) for an application) might be suitable. Given our specific relationship in (1), we consider the approach developed in Hansen and Seo (2002) where a two-regime threshold VECM model is proposed as a convenient method to combine nonlinearity and cointegration. In particular, the model allows for nonlinear adjustment to the long-run equilibrium. In fact, although a linear VECM model assumes a constant adjustment speed towards a long-run equilibrium, the threshold cointegration approach holds that error correction occurs depending on the threshold. Hansen and Seo (2002) define a two-regime nonlinear VECM of order $\ell + 1$ as follows:

$$\Delta \mathbf{x}_t = \begin{cases} \mathbf{A}'_1 \mathbf{X}_{t-1}(\boldsymbol{\beta}) + \mathbf{u}_t & \text{if } w_{t-1}(\boldsymbol{\beta}) \leq \gamma \\ \mathbf{A}'_2 \mathbf{X}_{t-1}(\boldsymbol{\beta}) + \mathbf{u}_t & \text{if } w_{t-1}(\boldsymbol{\beta}) > \gamma \end{cases} \quad (2)$$

with

$$\mathbf{X}_{t-1}(\boldsymbol{\beta}) = \begin{cases} 1 \\ w_{t-1}(\boldsymbol{\beta}) \\ \Delta \mathbf{x}_{t-1} \\ \Delta \mathbf{x}_{t-2} \\ \vdots \\ \Delta \mathbf{x}_{t-\ell} \end{cases} \quad (3)$$

where \mathbf{x}_t is a p -dimensional $I(1)$ time series with one $p \times 1$ cointegrating vector $\boldsymbol{\beta}$, $w_t(\boldsymbol{\beta}) = \boldsymbol{\beta}' \mathbf{x}_t$ denotes the $I(0)$ error correction term. In our case, with the previous notation, the vector of interest will be $\mathbf{x}_t = (g_t \ y_t)'$. The error \mathbf{u}_t is assumed to be a vector martingale difference sequence with finite covariance matrix $\boldsymbol{\Sigma} = E(\mathbf{u}_t \mathbf{u}_t')$. Then, γ is the threshold parameter and the coefficient matrices \mathbf{A}_1 and \mathbf{A}_2 govern the dynamics in the two regimes. The threshold effect has content if

$$\pi_0 \leq P(w_{t-1} \leq \gamma) \leq 1 - \pi_0 \quad (4)$$

where $\pi_0 > 0$ is a trimming parameter, which is set to be 0.05. Estimation of this model is performed using Maximum Likelihood (MLE) under the assumption that the errors are *iid* Gaussian.

Under this framework, we apply the test proposed by the same authors to verify the presence of threshold. Particularly, we test a null of linear cointegration (VECM) versus an alternative of threshold cointegration (two-regime VECM). It is a LM test of the following form:

$$\text{SupLM} = \sup_{\gamma_L \leq \gamma \leq \gamma_U} \text{LM}(\tilde{\beta}, \gamma) \quad (5)$$

where $\tilde{\beta}$ is the null estimate of β , the search region $[\gamma_L, \gamma_U]$ is set so that γ_L is the π_0 percentile of \tilde{w}_{t-1} and γ_U is the $(1 - \pi_0)$ percentile. This imposes constraint (4). The SupLM statistic has a nonstandard asymptotic distribution. As discussed in Hansen and Seo (2002), the fixed regressor bootstrap of Hansen (1996, 2000) can be used to calculate asymptotic critical values and p -values as well as residual bootstrap technique. This framework seems to be the correct one to consider potential nonlinear effects between our variables of interest if statistical testing produces such evidence.

3. Data

We use annual time series data of Italy's central government spending that comprise the years 1862-2009. Data on total government expenditure and its national defence item are at current prices and are drawn from the series that have recently been provided by Italy's State General Accounting Department (Ragioneria Generale dello Stato 2011). Spending refers to the total payments disbursed in the year, which have been obtained from the final budget of the state². Our data differ from those of the two most recent papers that test WL for Italy. In fact, Magazzino (2012) relies on the Informative Public Base (IBP), a database developed by the Bank of Italy that covers the shorter 1960-2008 time span and refers to the expenditure of the Italian public administration as a whole, i.e., it includes not just the expenditure of Italy's central government but also the expenditure of local governments (regional, provincial and municipal administrations). Our series differ also from those of Kuckuck (2014) who uses data drawn from Mitchell (2007) for the years 1850-1995 and from Eurostat for the years 1996-2010. These data are provided by Italy's State General Accounting Department, but, differently from ours, they refer to the expenses accrued and not to the actual payments in the fiscal year. For GDP at current prices and real per capita GDP (2010 prices) we rely on the new series of Italy's national accounts that have been recently provided by the Bank of Italy (Baffigi (2011)).

²From 1884 to 1964 Italy's fiscal year ran from July 1st to June 30th. Data have been attributed to the solar years by adding half of the expenditures disbursed in two consecutive fiscal years and assuming an equidistribution of the expenditure over each fiscal year.

Population data are from the *Ricostruzione della popolazione residente e del bilancio demografico* database (Istat (2012)).

Figure 1 depicts the evolution of the ratio of total government expenditure to GDP compared to real per capita GDP on the full sample³. It shows that between 1862 and the mid-1890s, the total government spending and real per-capita GDP show very similar trends. During this period the investments in the railways was particularly significant and spending for education and culture also increased constantly from Unification to the WWI. In the first fifty years following Unification, Italy's total government spending in real terms increased slowly, on average: between 1862 and 1913 total government spending was around 10% of GDP, but the participation in WWI led to a drastic increase of total government spending to the 35% of the GDP. In the years after the conflict, government spending sharply dropped again, in 1926 settling to the pre-war values compared to the GDP: defence spending dropped, while investments in public works and other economic interventions resumed. Participation in WWII led to another drastic increase of government spending, from just over 10% to more than 45% of the GDP. After the defeat of WWII, Italy was prohibited from reconstructing its own independent military power. This led to a drastic reduction in defence spending, as can be seen from Figure 2. On the other hand, economic and social components of expenditure started to increase: infrastructure, welfare, and redistribution by the state. Under pressure from the expansion of suffrage (universal suffrage was introduced in 1946), and, from the end of the 1960s, of an unprecedented wave of social conflicts, a progressive expansion of welfare services to new social categories took place in Italy until a universalistic welfare system was introduced in 1978. Moreover, Italy's pro-American stance during the Cold War and the possibility that the Italian Communist Party (the largest Communist party in the Western world) might organise a revolution meant that a large proportion of government expenditure was allocated to national security. In the 1970s, this expenditure was further increased to tackle political terrorism. Hence, between the end of WWII and 1963, public spending remained well below 25% of GDP while from the mid-1970s state spending began to grow more rapidly than the GDP, reaching a maximum peak of 44% in 1986 (see Figure 1). During recent years, the expenditure share was more stable. Between 1980 and 2009, total government expenditure actually grew less than proportionately to GDP. In particular, from 1993, with a view to Italy's signing up to the single European currency,

³See Tanzi and Schuknecht (2000) for a discussion of the role of government spending of the main industrialised economies in the 20th century and Cohen and Federico (2001) for an in-depth discussion of the Italian case.

the imbalance of the national accounts began to be tackled, and clear results were seen in 1995, making decisive progress in 1997, when the deficit dropped. The preliminary analysis of the series suggests a long-run relationship between total government spending and national income with some large deviations during the WWI and WWII period. This reinforces our idea to take into account some large but transitory deviations from the common long-run path of the variables.

4. Empirical Results

In this Section we explore the data to hand. Firstly, we conduct some preliminary investigation on stationarity of the series and existence of cointegrating relationship. Then, we test for the presence of nonlinearities in the data and proceed with estimation of the parametric model.

4a. Preliminary Analysis

To investigate stationarity of the two series of interest, we apply alternative unit root tests. The first test is the standard augmented Dickey-Fuller test which considers a null hypothesis of a unit root against the alternative of stationarity. The second is the modification of the above test proposed by Elliott (1999). The last is the KPSS test proposed by Kwiatkowski *et al.* (1992) for the null hypothesis of stationarity versus the alternative of non-stationarity. The results are reported in Table 1. Both tests lead us to consider government expenditure and real per capita GDP as realizations of $I(1)$ processes. Thus in the subsequent analysis we will consider the first difference of both series, and preliminary summary statistics are given in Table 2. Normality tests reject the null of normality in the data, and this could be partially due to temporal dependence in the moments of the series or to the presence of nonlinearities in the data.

Before proceeding further with the analysis, we search for the existence of a cointegrating relationship between the two variables of interest. Note that, as a preliminary condition for cointegration, we have already checked that the two time series are integrated of the same order ($I(1)$). Next, we implement the linear Johansen cointegration rank tests (Johansen (1988)), using 2 lags in the VAR, as suggested by BIC criterion, and including an unrestricted constant. As shown in Table 3, these preliminary tests easily reject the null hypothesis of no-cointegration, indicating the presence of one cointegrating relation. Therefore, we estimate the linear VECM, and the results are shown in Table 4. Here the estimates indicate bidirectionality in the two variables and this shows no evidence of the WL for this country. However, this could be due to a misspecification of the correct model owing to the threshold-type nonlinearities present in the data. That is why we apply a threshold

cointegration technique in what follows.

4b. Threshold Cointegration

Firstly, we test explicitly for the presence of threshold effect under the null hypothesis of linear cointegration for the complete bivariate specification. The fixed regressor bootstrap and the residual bootstrap methods were used, and both were simulated using 5,000 bootstrap replications. A lag length of $\ell = 1$ in Equation (3) was selected, basing our choice on AIC and BIC criteria. Table 5 reports the test results for the linear versus nonlinear cointegration hypothesis, together with threshold and cointegrating parameter estimates. The resulting LM statistic computed as a function of the threshold parameter estimate γ is plotted in Figure 3. The fixed regressor bootstrap method reject the null at 5% confidence level and the residual bootstrap method at 1%. Thus, the threshold cointegration model seems more appropriate to our data than a linear one. In fact, neglecting the asymmetric adjustment may lead to biased inferences and misleading conclusions. Hence the error correction mechanism differs depending on deviations from the equilibrium below or above the threshold parameter. So that, the first regime (say, "normal") corresponds to $g_{t-1} - 0.501y_{t-1} \leq -5.532$ and the second regime (say, "extreme") corresponds to $g_{t-1} - 0.501y_{t-1} > -5.532$. We also observe that 87% of all the observations belong to the first regime and the remaining 13% to the second regime. Estimation is performed by MLE following the grid-search algorithm proposed by Hansen and Seo (2002) over a 80×80 grid on the parameters β and γ . Table 6 reports estimated values of the threshold VECM. There are three most significant results. The first is that in "normal" times lagged values of real per capita GDP significantly influence the dynamic behavior of the Italian economy, while in the "extreme" regime government expenditure of the previous period tends to matter most. The second result concerns the coefficients of error-correction term w_{t-1} in the two regimes. This term strongly suggests causality running from economic growth to government activity. In fact, the negative and statistically significant adjustment parameters in both government equations imply validity of WL as GDP being a driving force of government expenditure. Note also that the magnitude of the response of government expenditure is between 9 ("normal" regime) and 5 ("extreme" regime) times greater than the coefficient in the GDP equation. Finally, the last outcome concerns the estimated long-run relationship between government expenditure and national income. In fact it results significantly greater than zero (p -value=0.009) suggesting that government expenditure is income elastic, i.e a superior good, over the full sample. In conclusion, all these results validate WL for the Italian economy from 1862 to 2009. To allow visual interpretation of the results, the error

correction mechanism is pictured in Figure 4. It can be noted that the strong error-correction effect for the two variables is depicted on the right-hand side of the estimated threshold. On the contrary, it shows a flat near-zero effect for real per capita GDP and slightly greater effect for government expenditure on the left-hand side. Asymmetry shows a stronger error-correction effect in the "extreme" regime compared to the "normal" regime which is driven by the public side of the economy. This finding seems to provide new evidence of nonlinearities in the underlying parametric processes. Moreover, the diagnostics reported at the bottom of Table 5 reinforce the evidence of nonlinearity given that the null of equality of the dynamic coefficients as well as equality of coefficients in the error-correction term are strongly rejected. Furthermore, we investigate the timing of the regime shift in the cointegrating relationship. The estimated regime classification according to the threshold VECM is visually presented in Figure 5. Our estimated model suggests that the economy happens to be in the "extreme" regime between 1915 - 1919 (with some follow-up in the 1921 - 1923) and between 1934 - 1945. These facts are in line with the preliminary inspection conducted in Figure 2, pointing out a drastic increase in military spending in those time spans.

5. Robustness

As discussed before, during the process of economic development of the Italian economy there is a tendency of government expenditure to grow relative to GDP according to WL. However, this evidence clearly emerges only when we take into account nonlinear adjustment of the variables to their long-run growth path during the WWI and WWII period. In this Section, we conduct two checks for correctness of our outcome. Firstly, we verify the linearity of WL of public spending when the "abnormal" periods are excluded by the data. Secondly, we test the linearity of this relationship subtracting the military spending from the total expenditure series. Particularly, we run the linearity versus nonlinearity LM test on the two subsamples left when we exclude periods which are sources of nonlinearity. Once we run the fixed regressor and residual bootstrap for the subsample 1862 - 1914 (1946 - 2009, respectively) we obtain p -values equal to 0.39 and 0.054 (0.092 and 0.052, respectively) which do not reject the null hypothesis of linearity at 5% significance level. Since the model especially captures the WWI and WWII, it suggests military disbursement as a driving force for the government expenditure. If this is the case, once we subtract public spending on defence from the total government expenditure, we expect to remove the source of nonlinearity in the data. Therefore, we construct a new variable as the natural logarithm of the ratio between total government expenditure minus national

defence expenditure over nominal GDP. We run again the linearity LM test for the bivariate model and the p -values from fixed regressor and residual bootstrap are 0.26 and 0.105, respectively. This means that we do not reject the linearity hypothesis at 10% significance level, suggesting exactly military disbursement as a driving force for the government expenditure. This evidence reinforces the conclusion outlined in Section 4. The model for the evolution of the demand of public services is basically linear according to WL. The source of non linearities is large but transitory and it is simply due to abnormal behaviour of the expenditure for national defence during the WWI and WWII period.

6. Conclusion

In this paper we test the long-run tendency for Italian total government expenditure to grow relative to per capita GDP over the period 1862-2009. We find evidence of a threshold cointegrating relationship between these variables which turns out to be consistent with WL, given the different adjustment speeds to their long-run growth path. Asymmetric error-correction effects identify two different regimes and the WWI and WWII years completely describe one of them. The abnormal response of government spending in the latter regime is only due to the increase in defence expenditure during the war periods. On the basis of these results, we conclude that the model for public spending is basically linear and consistent with an expanding government sector as the economy progresses.

Acknowledgements. The authors gratefully acknowledge financial support from PRIN-MIUR Grant 2010J3LZEN-003 and the research grant FAR 2014 (provided by University of Modena and Reggio E., Italy).

References

- [1] Atkinson, A.B., and Stiglitz, J.E. (1980) *Lectures on public economics*, MacGraw-Hill, New York.
- [2] Baffigi, A. (2011) *Italian national accounts, 1861 - 2011*, Quaderni di Storia Economica. Economic History Working Papers No. 18, Banca d'Italia, Roma.
- [3] Baumol, W.J. (1967) *Macroeconomics of unbalanced growth: the anatomy of urban crisis*, The American Economic Review **57**(3), 415-426.
- [4] Bohl, M.T. (1996) *Some international evidence on Wagner's law*, Public Finance / Finances Publiques **5**, 185-200.
- [5] Cohen, J., and Federico, G. (2001) *The development of the Italian economy, 1820-1960*, Cambridge University Press, Cambridge.
- [6] Durevall, D., and Henrekson, M. (2011) *The futile quest for a grand explanation of long-run government expenditure*, Journal of Public Economics **95**(7), 708-722.
- [7] Elliott, G. (1999) *Efficient tests for a unit root when the initial observation is drawn from its unconditional distribution*, International Economic Review **40**(3), 767-784.
- [8] Ferraresi, T., Roventini, A., and Fagiolo, G. (2015) *Fiscal policies and credit regimes: a TVAR approach*, Journal of Applied Econometrics **30**, 1047-1072.
- [9] Ghate, C., and Zak, P.J. (2002) *Growth of government and the politics of fiscal policy*, Structural Change and Economic Dynamics **13**, 435-455.
- [10] Gregory, A.W., and Hansen, B.E. (1996) *Residual-based tests for cointegration in models with regime shifts*, Journal of Econometrics **70**(1), 99-126.
- [11] Hansen, B.E. (1996) *Inference when a nuisance parameter is not identified under the null hypothesis*, Econometrica **64**, 413-430.

- [12] Hansen, B.E. (2000) *Testing for structural change in conditional models*, Journal of Econometrics **97**(1), 93-115.
- [13] Hansen, B.E., and Seo, B. (2002) *Testing for two-regime threshold cointegration in vector error-correction models*, Journal of Econometrics **110**(2), 293-318.
- [14] Henrekson, M. (1993) *Wagner's law - a spurious relationship?*, Public Finance / Finances Publiques **48**, 406-415.
- [15] Istat (2012), "Ricostruzione della popolazione residente e del bilancio demografico" (https://www.seriestoriche.istat.it/Tavola_2.3.1.xls).
- [16] Johansen, S. (1988) *Statistical analysis of cointegration vectors*, Journal of Economic Dynamics and Control **12**(2), 231-254.
- [17] Kuckuck, J. (2014) *Testing Wagner's law at different stages of economic development*, FinanzArchiv: Public Finance Analysis **70**(1), 128-168.
- [18] Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. (1992) *Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?*, Journal of Econometrics **54**(1), 159-178.
- [19] Magazzino, C. (2012) *Wagner versus Keynes: public spending and national income in Italy*, Journal of Policy Modeling **34**, 890-905.
- [20] Mitchell, B.R. (2007) *International Historical Statistics 1750 - 2005*, Vol. 3, 6th ed. Palgrave Macmillan, New York.
- [21] Oxley, L. (1994) *Cointegration, causality and Wagner's law: a test for Britain 1870 - 1913*, Scottish Journal of Political Economy **41**, 286-298.
- [22] Tanzi, V., and Schuknecht, L. (2000) *Public spending in the 20th century*, Cambridge University Press, Cambridge.
- [23] Thornton, J. (1999) *Cointegration, causality and Wagner's law in 19th century Europe*, Applied Economics Letters **6**, 413-416.
- [24] Tsay, R.S. (1998) *Testing and modeling multivariate threshold models*, Journal of the American Statistical Association **93**, 1188-1202.
- [25] Wagner A. (1883) *Finanzwissenschaft*, 3rd edition, Leipzig. Partly reprinted in *Classics in the Theory of Public Finance*, edited by R.A. Musgrave and A.T. Peacock (1958), Macmillan, London.

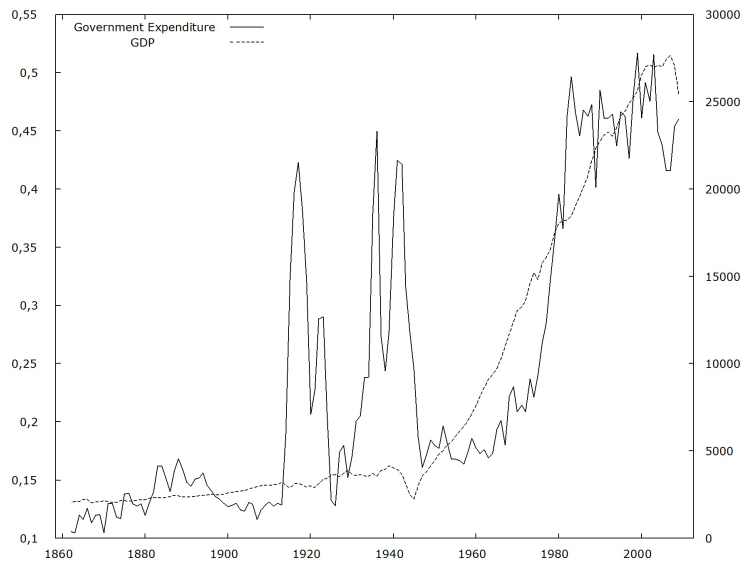


Figure 1: Total government expenditure as a share of GDP and real per capita GDP (1862-2009).

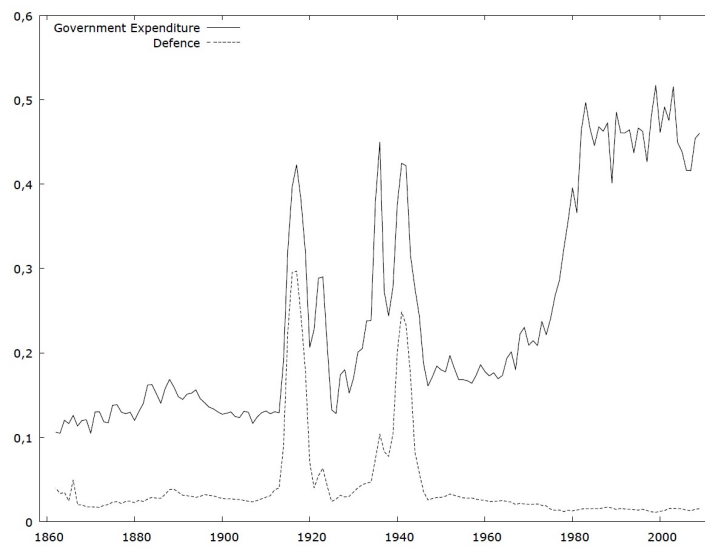


Figure 2: Total government expenditure as a share of GDP and national defence expenditure as a share of GDP (1862-2009).

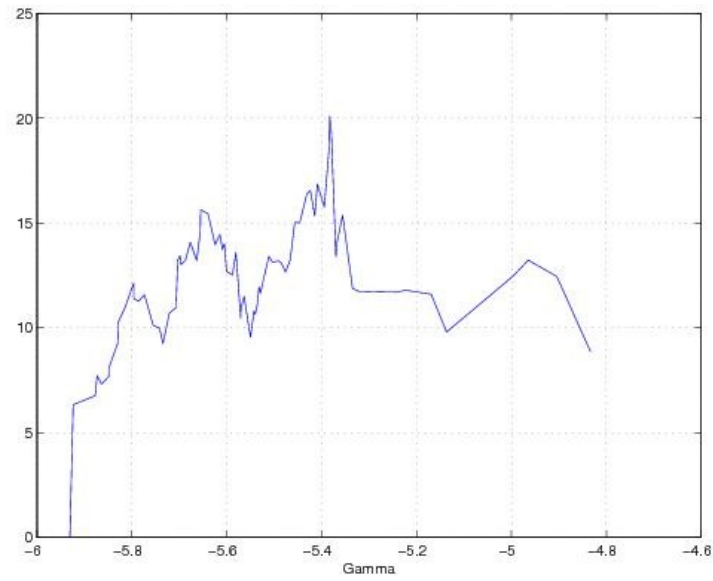


Figure 3: Lagrange Multiplier (LM) statistic for the bivariate (government expenditure and real per capita GDP) threshold cointegration model as a function of the threshold parameter γ .

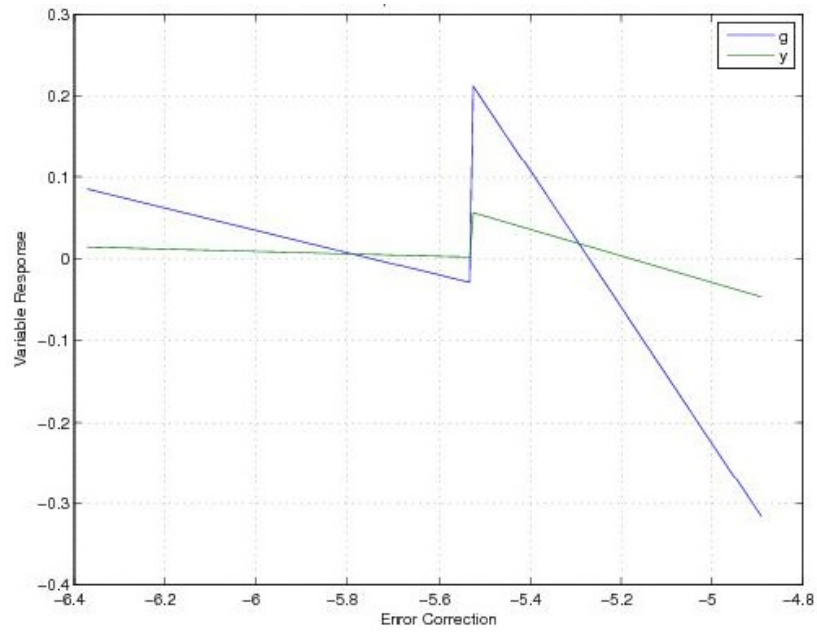


Figure 4: Variable response to error-correction. Variables are government expenditure (g) and real per capita GDP (y).

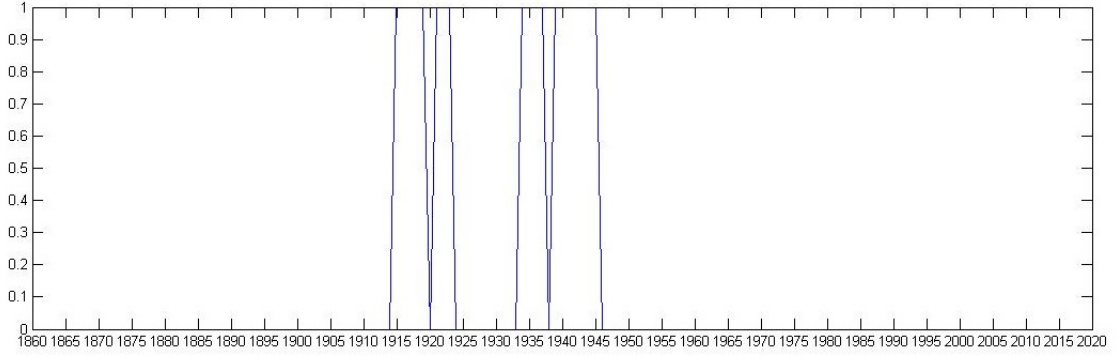


Figure 5: Timing of the realization of "extreme" regime obtained from the threshold VECM in government expenditure (g) and real per capita GDP (y).

<i>Series</i>	ADF t -tests		ADF-GLS tests		KPSS tests	
	t_μ	t_τ	t_μ	t_τ	η_μ	η_τ
g	-1.541 [2]	-2.977 [2]	-0.511 [2]	-3.002 [2]	3.349***[2]	0.206** [2]
Δg	-9.809***[1]	-9.775***[1]	-9.405***[1]	-9.471***[1]	0.025 [1]	0.024 [1]
y	0.443 [2]	-1.785 [2]	1.479 [2]	-1.091 [2]	6.831***[2]	1.046***[2]
Δy	-7.374***[1]	-7.461***[1]	-7.653***[1]	-7.485***[1]	0.372 [1]	0.127 [1]

Table 1: Unit root and stationary tests for total government expenditure (denoted by g) and real per capita GDP (denoted by y). The symbol Δ denotes the first-difference transformation of the series. The symbol *** indicates significance at the 99% confidence level, while the symbol ** indicates significance at the 95% confidence level. Those symbols refer to the choice of not rejecting the null hypothesis both for ADF (presence of unit root) and KPSS (stationarity) tests. Numbers in square brackets refer to the selected lag order for each statistics. Finally, t_μ and t_τ correspond to test statistics where the auxiliary regression contains a constant and a trend, respectively; η_μ and η_τ are test statistics for level and trend stationary, respectively.

<i>Statistics</i>	Δg_t	Δy_t
Mean	0.010	0.017
Standard deviation	0.139	0.049
Skewness	-0.088	0.106
Kurtosis	3.645	10.118
$J - B$	81.551***	627.41***
$S - W$	0.922***	0.840***

Table 2: Summary statistics on first-differenced series of total government expenditure (g) and real per capita GDP (y). $J - B$ and $S - W$ denote the Jarque-Bera and Shapiro-Wilk tests for the null of normality, respectively. The symbol *** indicates significance of 1% significance level

<i>Series</i>	g_t	y_t	
<i>Trace test</i>			
No. of CE(s)	Eigenvalue	Trace Statistic	p -value
None	0.1269	19.830	0.0092***
At most 1	0.00006	0.0091	0.9240
<i>Maximum Eigenvalue test</i>			
No. of CE(s)	Eigenvalue	Max-Eigen Statistic	p -value
None	0.1269	19.821	0.0048***
At most 1	0.00006	0.0091	0.9240

Table 3: Johansen cointegration rank tests (CE stands for Cointegrating Equation) based on VAR(2) with unrestricted constant. The symbol *** denotes rejection of the hypothesis at the 0.01 level. Both trace test and max-eigenvalue tests indicate 1 cointegrating equations at 0.01 level.

<i>Variables</i>	Δg_t	Δy_t
Cointegrating vector	1.0000 (0.0000)	-0.5124 (0.0718)
Intercept	-0.6953*** (0.2039)	-0.1909** (0.0741)
w_{t-1}	-0.1186*** (0.0342)	-0.0349*** (0.0124)

Table 4: Estimates of the linear VECM(1) with unrestricted constant for government expenditure (g) and real per capita GDP (y). Standard errors are provided in parentheses. The symbol *** denotes significance at 99% significance level and the symbol ** at 95% significance level.

Estimates	
Threshold parameter estimate (γ)	-5.532
Cointegrating parameter estimate (β)	0.501
<i>Lagrange multiplier threshold test</i>	
sup LM value	20.082
p -value of fixed regressor bootstrap	0.024
p -value of residual bootstrap	0.009

Table 5: Lagrange Multiplier (LM) test for threshold cointegration between government expenditure and real per capita GDP. The number of gridpoints for threshold and cointegrating vector is equal to 80. For p -values, the number of bootstrap replications is set to 5000.

<i>Variables</i>	1st regime (87% obs)		2nd regime (13% obs)	
	Δg_t	Δy_t	Δg_t	Δy_t
Intercept	-0.136** (0.055)	-0.015 (0.014)	-0.834*** (0.219)	-0.162 (0.089)
w_{t-1}	-0.781** (0.334)	-0.079 (0.081)	-4.395*** (1.143)	-0.839 (0.458)
Δg_{t-1}	0.105 (0.154)	0.029 (0.028)	0.631*** (0.094)	0.086 (0.061)
Δy_{t-1}	-0.350 (0.198)	0.495*** (0.074)	-0.019 (0.433)	0.179 (0.307)
Wald test equality dynamic coefs.	19.661	$(p\text{-value: } 0.001)$		
Wald test equality EC coefs.	18.923	$(p\text{-value: } 0.001)$		

Table 6: Estimates of the threshold VECM for government expenditure (g) and real per capita GDP (y). Eicker-White standard errors are provided in parentheses. In Wald test diagnostics, the null hypothesis is equality of the dynamic coefficients and of the coefficients on the error correction terms across the two regimes, respectively. The symbol *** denotes significance at 99% significance level and the symbol ** at 95% significance level.